WHAT IS CLAIMED:

1. A method for training a kernel-based learning machine using a dataset comprising:

filling a kernel matrix with a plurality of kernels, each kernel comprising a pairwise similarity between a pair of data points within a plurality of data points in the dataset;

defining a fully-connected graph comprising a plurality of nodes and a plurality of edges connecting at least a portion of the plurality of nodes with other nodes of the plurality, each edge of the plurality of edges having a weight equal to the kernel between a corresponding pair of data points, wherein the graph has an adjacency matrix that is equivalent to the kernel matrix;

computing a plurality of eigenvalues for the kernel matrix;

selecting an eigenvector corresponding to the smallest non-zero eigenvalue of the plurality of eigenvalues;

bisecting the dataset using the selected eigenvector; and

training the kernel-based learning machine using at least a portion of the bisected dataset.

2. The method of claim 1, further comprising, after computing a plurality of eigenvalues, determining a number of clusters of data points within the dataset by identifying all zero eigenvalues.

3. The method of claim 1, further comprising:

computing a second eigenvector; and

minimizing a cut cost for bisecting the dataset by applying a threshold to the second eigenvector.

4. The method of claim 3, wherein the threshold limits the second eigenvector to entries of -1 and +1.

5. The method of claim 1, wherein the data points within the dataset are unlabeled and the step of bisecting the dataset comprises assigning the data points to a cluster of a plurality of clusters.

6. The method of claim 1, wherein the data points within a first portion of the dataset are labeled and the data points of a second portion of the dataset are unlabeled, and wherein the step of filling the kernel matrix comprises:

selecting a kernel K;

normalizing the selected kernel K to $-1 \leq K \leq +1$; and

if both data points of a pair come from the first portion of the dataset, the corresponding kernel comprises a labels vector.

7. The method of claim 6, further comprising:

calculating a second eigenvector of the kernel matrix to obtain an alignment;

thresholding the second eigenvector; and

based on the alignment, assigning labels to the unlabeled data points.

8. The method of claim 7, further comprising adjusting at least a portion of the plurality of kernels to align the second eigenvector with a pre-determined label.

9. The method of claim 1, further comprising, prior to computing a plurality of eigenvalues, computing a first eigenvector and assigning a rank to each of the plurality of data points based on popularity.

10. The method of claim 9, further comprising identifying as dirty any data points of the plurality having a low rank.

11. The method of claim 10, further comprising cleaning the dirty data points.

12. A spectral kernel machine comprising:

at least one kernel selected from a plurality of kernels for mapping data into a feature space, the at least one kernel selected by training the plurality of kernels on a dataset comprising a plurality of data points wherein the dataset is divided into a plurality of clusters by applying spectral graph theory to the dataset and selecting the at least one kernel that is optimally aligned with the division between the plurality of clusters.

13. The spectral kernel machine of claim 12, wherein the division between the plurality of clusters is determined by a first eigenvector in an adjacency matrix corresponding to a graph comprising a plurality of nodes comprising the plurality of data points.

14. The spectral kernel machine of claim 12, wherein the dataset is unlabeled.

15. The spectral kernel machine of claim 12, wherein the dataset is

partially labeled.

16. A spectral kernel machine comprising:

at least one kernel selected from a plurality of kernels for mapping data into a feature space, the at least one kernel selected by training the plurality of kernels on a dataset comprising a plurality of data points wherein the dataset is bisected into a plurality of clusters by applying spectral graph theory to the dataset and selecting the at least one kernel that minimizes a cut cost in the dichotomy between the plurality of clusters.

17. The spectral kernel machine of claim 16, wherein the dichotomy between the plurality of clusters is determined by a first eigenvector in an adjacency matrix corresponding to a graph comprising a plurality of nodes comprising the plurality of data points.